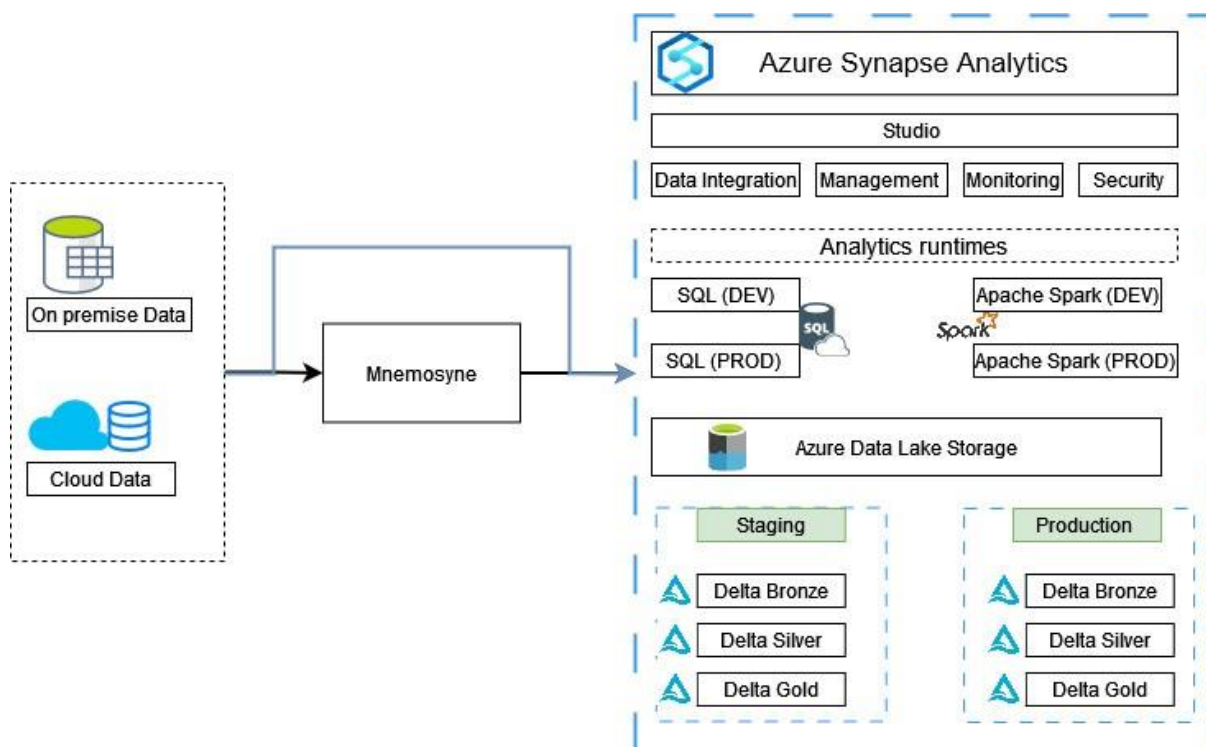


ΔΟΜΗ ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΕΡΓΑΣΙΑΣ / ΑΡΧΙΤΕΚΤΟΝΙΚΗ

Σύμφωνα να την κατανόησή μας από τις περιγραφόμενες ανάγκες και εσωτερικές ροές της ΔΕΗ στο έντυπο προδιαγραφών του έργου, παραθέτουμε το flow chart της προσέγγισης μας



Το Azure Synapse Analytics είναι μια ισχυρή και ενοποιημένη υπηρεσία ανάλυσης δεδομένων που παρέχεται στο Microsoft Azure. Συνδυάζει την επεξεργασία μεγάλων δεδομένων και τις δυνατότητες αποθήκευσης δεδομένων σε ένα μόνο πακέτο λύσης, επιτρέποντας στις οργανώσεις να αναλύουν μεγάλους όγκους δεδομένων αποδοτικά και να αποκτήσουν πολύτιμες εισηγήσεις. Η αρχιτεκτονική της ανάλυσης δεδομένων στο Azure Synapse περιλαμβάνει αρκετά βασικά στοιχεία που συνεργάζονται για την παροχή ολοκληρωμένων δυνατοτήτων ανάλυσης.

- **Εισαγωγή Δεδομένων:** Η διαδικασία ανάλυσης δεδομένων ξεκινά με την εισαγωγή δεδομένων από διάφορες πηγές στο Azure Synapse. Αυτό μπορεί να περιλαμβάνει δομημένα, ημιδομημένα και αδομητά δεδομένα από βάσεις δεδομένων, πλατφόρμες και άλλες εξωτερικές πηγές. Το Azure Synapse υποστηρίζει μια ευρεία γκάμα μεθόδων εισαγωγής, όπως το Azure Data Factory, η αποθήκευση του Azure Data Lake, η αποθήκευση του Azure Blob κ.λπ. Τα δεδομένα μεταφέρονται στον χώρο εργασίας Synapse, όπου μπορούν να προετοιμαστούν για περαιτέρω επεξεργασία.
- **Προετοιμασία και Μετασχηματισμός Δεδομένων:** Μόλις τα δεδομένα εισαχθούν στο Azure Synapse, υποβάλλονται σε βήματα προετοιμασίας και μετασχηματισμού για να γίνουν

κατάλληλα για ανάλυση. Το Azure Synapse παρέχει ισχυρές δυνατότητες ενσωμάτωσης δεδομένων μέσω εργαλείων όπως το Azure Data Factory, το Azure Databricks και οι SQL Server Integration Services (SSIS). Αυτά τα εργαλεία επιτρέπουν τον καθαρισμό των δεδομένων, την επεξεργασία τους και τον μετασχηματισμό τους, επιτρέποντας στις οργανώσεις να διαμορφώσουν τα δεδομένα σύμφωνα με τις αναλυτικές απαιτήσεις τους.

- **Αποθήκευση και Διαχείριση Δεδομένων:** Το Azure Synapse αξιοποιεί δύο βασικά στοιχεία για την αποθήκευση και τη διαχείριση δεδομένων: το Azure Data Lake Storage και οι κατανεμημένες αποθηκευτικές δυνατότητες βασισμένες σε SQL. Το Azure Data Lake Storage λειτουργεί ως ένα εξαιρετικά κλιμακούμενο και οικονομικό αποθηκευτικό χώρο για την αποθήκευση μεγάλων όγκων ακατέργαστων και επεξεργασμένων δεδομένων. Παρέχει έναν ιεραρχικό αρχαιοθέτη που μπορεί να χειριστεί διάφορους τύπους και μορφές δεδομένων. Από την άλλη πλευρά, οι κατανεμημένες αποθήκευσης δεδομένων βασισμένες σε SQL, γνωστές ως Azure Synapse SQL Pools, επιτρέπουν τη δημιουργία και διαχείριση αποθήκων δεδομένων που είναι βελτιστοποιημένες για αναλυτικές εργασίες. Παρέχει δυνατότητες ερωτήσεων υψηλής απόδοσης και υποστηρίζει τυπικές ερωτήσεις SQL.
- **Επεξεργασία Δεδομένων και Ανάλυση:** Το Azure Synapse προσφέρει πολλές επιλογές για την επεξεργασία δεδομένων και την ανάλυση, καλύπτοντας διάφορες χρήσεις και απαιτήσεις φορτίου εργασίας. Ένα από τα βασικά στοιχεία είναι το Apache Spark, που παρέχει δυνατότητες κατανεμημένης επεξεργασίας για εργασίες μεγάλων δεδομένων. Το Azure Synapse χρησιμοποιεί το Azure Databricks, μια πλατφόρμα ανάλυσης βασισμένη στο Apache Spark, για να επιτρέψει την κλιμακούμενη εξερεύνηση δεδομένων, την προετοιμασία δεδομένων και την προηγμένη ανάλυση. Επιπλέον, οι κατανεμημένες αποθήκες δεδομένων Azure Synapse SQL παρέχουν ένα βελτιστοποιημένο περιβάλλον για την εκτέλεση παραδοσιακών ερωτημάτων SQL σε δομημένα και ημιδομημένα δεδομένα.
- **Οργάνωση και Παρακολούθηση:** Το Azure Synapse ενσωματώνεται με το Azure Data Factory για να παρέχει δυνατότητες οργάνωσης δεδομένων. Το Azure Data Factory επιτρέπει τη δημιουργία ροών για τον προγραμματισμό και οργάνωση της εκτέλεσης διαφόρων εργασιών ενσωμάτωσης και μετασχηματισμού δεδομένων. Παρέχει ένα οπτικό περιβάλλον για τον σχεδιασμό, την παρακολούθηση και τη διαχείριση των αγωγών δεδομένων, εξασφαλίζοντας την ομαλή ροή των δεδομένων κατά τη διάρκεια της αναλυτικής διαδικασίας. Το Azure Synapse περιλαμβάνει επίσης χαρακτηριστικά παρακολούθησης και διαχείρισης που επιτρέπουν στις οργανώσεις να παρακολουθούν την απόδοση, την κλιμάκωση και την υγεία των διαδικασιών ανάλυσης δεδομένων.

Η προτεινόμενη αρχιτεκτονική της ανάλυσης δεδομένων στο Azure Synapse παρέχει ένα ισχυρό περιβάλλον για την ανάπτυξη και εκτέλεση αναλυτικών εργασιών σε μεγάλη κλίμακα. Η τελική αρχιτεκτονική της λύσης θα καταγραφεί με περισσότερη λεπτομέρεια κατά τη διάρκεια της πρώτης εβδομάδας και αφού πρώτα έχουμε την πλήρη εικόνα της υφιστάμενης λύσης/υποδομής.

